



**AIRA**  
AMERICAN IMMUNIZATION  
REGISTRY ASSOCIATION

## IIS Data Cleansing

---

**Takeaways from the IIS Data Analyst  
Collaborative (IDAC)**

May 2025

# IIS Data Analyst Collaborative (IDAC)

May 15, 2025

## The IIS Data Analyst Collaborative (IDAC)

The IIS Data Analyst Collaborative (IDAC) is a discussion-based collaborative for people who work with immunization data to connect on important and timely topics. IDAC happens quarterly on the third Thursday at 1 p.m. ET. There will be a different discussion topic every meeting. The goals for these collaboratives are to offer an opportunity to share and develop skills and insights, to help people who work with immunization data to connect and spark collaborations, and to create a supportive, engaged community of IIS data analysts.

## IIS Data Cleansing

Questions discussed during this IDAC session included the following:

- What are issues with merging records, and what are some solutions?
- How can demographics be improved?
- What data cleansing steps do you take outside of the IIS prior to data analysis?
- Any other issues with data cleansing?

### Discussion takeaways

#### *What are issues with merging records, and what are some solutions?*

- It requires dedicated staff and is an ongoing, large job.
- An IIS does not have a way of determining how big of a problem it has.
- IIS are starting to see some data quality issues from newer providers/COVID providers.
- Texas's IIS presented during the 2025 AIRA National Meeting (presentations will be posted to the Repository soon, and AIRA staff will link to this presentation on the forum): "Duplicate Patients via Data Exchange: Strategic Remediation on Core Issues," which describes the areas where duplicates are added to IIS and how to prevent them. This will help to save the time and effort of merging records later.
  - Texas explained the enhancements it made to its IIS application to help reduce the number of duplicates in the system. The enhancement falls on the payor side and helps prevent the creation of duplicates on the way into the IIS.

#### *Baby names*

- There are often baby name issues where birth clerks are under pressure to report the birth doses within 24 to 48 hours, but the baby is not named that quickly.

- If an IIS limits the usage of “baby boy” or “baby girl,” then the facilities get creative with their naming thus making the records harder to locate.
- Baby name solutions
  - Do not accept any name with “baby” in the field (reject via HL7).
    - Reject the whole message if the name is on the JUNK\_NAME table. Periodically update that table as providers come up with creative ways to circumvent it.
  - Work with Vital Records and the birthing facility so that the birth clerks enter the birth doses into the Vital Records system when entering the birth.
  - Work with the EMR vendor (example given was Epic) to resubmit the birth dose VXU messages at discharge when the record has been updated with the baby’s given name.
    - Get all EHRs to either rely only on the Vital Records or wait for discharge when they update to the legal name (Epic, ECW, Allscripts, etc.).
  - Getting the medical record number (MRN) from Vital Records birth reporting to the IIS can be used to match with the MRN from the unnamed baby record.
    - Collect birth file numbers that can be used for deduplication.
    - One IIS accepts baby names but requests the MRN to match it to the MRN received from Vital Statistics. That IIS also requires mother’s name to match mother’s information from Vital Records. It’s very lengthy, but the process has been successful.
  - Conduct regular training with birthing facilities regarding their records and work with them to wait until discharge to report the birth doses to the IIS.
  - Run regular weekly checks on birth hospital submissions to look for new data submissions.

#### *How can demographics be improved?*

- Run race and ethnicity completeness reports frequently and track issues with providers. Reach out to help them improve their data.
  - Several IIS conduct outreach performed by their data quality team and have seen an improvement in race and ethnicity submissions since the outreach began.
- Send out error reports to HL7 providers; include the actual error messages in the reports.
- Require eligibility and administered by completeness above 80%. One IIS has a data quality stats report that pulls up for each clinic and, if below 80%, includes this information in the email to both the EHR vendor and the clinic contact. If the provider does not improve, then its HL7 traffic is taken back to the test environment.
- One participant asked if anyone could share examples of data quality reports.
  - Kentucky and Connecticut are currently working with their vendors to create monthly automated data quality reports that the clinic staff will be able to access.

- One IIS also conducts outreach for any test patient data entered in the production environment. This has helped clean up the IIS database and merge some of the patient records (i.e., patients previously identified as John Doe).
- An issue was raised where the IIS continues to collect race and ethnicity and keeps any of the varying values submitted over time for each patient. Has any IIS worked to resolve this?
- One IIS sends data-quality follow-up reports to all providers that administered vaccines within the last month, including both dose data completeness and demographic data completeness. If any of the required data elements fall below 90% completeness, the IIS emails the results to the primary contacts at the provider sites. As a newer enhancement, the IIS has a missing-dose data report and a missing-demographic data report that its providers can run on their own for any administration date range. The querying providers can see any records missing required information. The IIS is currently building a data quality page on its home screen so, as soon as providers log in, they see the data completeness metrics for their patients for doses administered in the previous calendar month. There are automatic highlights for anything that does not meet the 90% goal. The IIS expects this tool to be in production soon.

*What data cleansing steps do you take outside of the IIS prior to data analysis?*

- Before running analysis, make sure the record has at least one immunization or is the birth record from Vital Stats to make sure it is a valid record.
- IIS expressed issues with deciding how to remove people who have moved (i.e., analysis to determine if someone who stopped getting immunizations still lives in the area or if that person has truly moved out of the area).
  - Oregon presented during the 2025 AIRA National Meeting (presentations will be posted to the Repository soon, and AIRA staff will link to this presentation on the forum). It was a four-state comparison of denominator inflation reduction methods. The main lesson is that there are several very viable choices, and IIS should choose one to do.
  - Other IIS also see the problem of becoming a non-vaccinator. When conducting analysis, be careful of what your purpose is. If needing to find pockets of need, then the county-level analysis would not give that level of detail. In these analyses, you would want to count those that have stopped vaccinating. Whereas, if you are calculating a vaccination rate by county, then selecting records that indicate a received vaccination in the last five years in that area/ZIP code would be sufficient.
  - IIS are experiencing more issues with anti-vaccinators in some areas. Inactivating some for not having a vaccine in a five-year period for analysis means we are taking these persons out even though our providers and county staff know they exist. We really struggle with finding out how to find the way to determine who really should not be counted and who really should be counted but needs outreach.
  - Analyses are questioned whether the data is missing or inflated. The ability to track people over time comes into play. If you need to know on a local level where people are right now relative to the shots they received years ago, this is a challenge for everyone.

- There can be an inflation issue with older people, as seniors can bounce between facility types. What are people doing to reduce for both mobility and death? Some IIS receive death certificates from Vital Statistics. It is a flat file and might be able to be processed to match on FN, LN, DOB, and SSN.
- One IIS assesses the immunization rate by county and checks the city-county agreement that allows it to correct the city values' spelling and match the record with the correct county.
- With COVID-19 immunizations, an IIS checks if the CVX codes match with patient age.

#### *Any other issues with data cleansing?*

- Is there a sweet spot for inactivating patients (e.g., after a certain age like 5 years, 7 to 10 years)?
  - One IIS staff works with large provider sites and goes through patient lists (5 to 7 years depending on age group).
  - Another IIS uses school enrollment information to review if the child is enrolled in the school system before inactivating.
  - It is trickier with adults. But, with HL7 reporting, it is possible to assume that, if an adult has not received any shots in more than 10 years, that adult can be inactivated. Many IIS have business rules that will reactivate patients if they receive a new immunization.
  - Inactivation is nuanced! There are two levels:
    - 1) Excluding a patient from population analysis if he or she has not received an immunization in five years
    - 2) Permanent inactivation
- Is there any consideration or evidence as to what should be done for these scenarios?
  - One IIS referenced the [Management of Patient Status in the IIS MIROW guide](#), which provides some guidance on patient inactivation at provider level and population level.